Test cases

You must run the agent manually for the following test cases, note the outputs, and then add them to the evaluation set.

Valid – No Escalation Required

1. Name: Mihai
   Question: Can I get reimbursed for a taxi ride I took to a client meeting?
   → Agent should mention that taxis for client meetings are reimbursable and reference the policy briefly.

2. Name: Andreea
   Question: How do I request vacation time? Is there a form or process I should follow?
   → Agent should describe the standard leave request process and where the form or system can be accessed.

3. Name: Dan
   Question: I forgot to clock out yesterday — what should I do now?
   → Agent should explain the time correction process and whom to contact if needed.

4. Name: Vlad
   Question: Hi, I'm a new hire — just started this week. What steps should I follow to get onboarded properly?
   → Agent should guide through the onboarding steps and mention any checklist or portal.

5. Name: Cristina
   Question: I'm feeling unwell today. How do I officially notify the company about a sick leave?
   → Agent should detail the notification process and if any form or medical note is required.

Evaluation

After receiving answers from the agent:

1. Add each test case and response to the evaluation set.

2. Run the evaluation.

   a. If the evaluation score is below 90%, review the answers manually.

   b. If the responses are semantically correct but phrased differently, you must tune the evaluator prompt accordingly.

3. Adjust the evaluator prompt until your score reaches 95%+ for meaningful, valid answers.